The Story of High-Achievers

Ameneh Fadaie afadaie@mit.edu

Austin M. Parker amparker@mit.edu

Lisa Baer Lisa.baer@mit.edu

ABSTRACT

Through our visualization, we wanted to explore the common environment and features that facilitate scientific creativity and increase the academic achievements of researchers. The best examples to illustrate this is by looking at Nobel laureates to see how and in what conditions they won their awards. We visualize information about the laureate fields, the age they won the prize, the number of publications they have, whether they stayed in their born country or they moved to other countries, and whether they collaborated with other researchers or if they worked alone. We illustrate this data in the shape of a flower and each flower represents one laureate. We used scrolly telling to teach our encodings to our users step by step, and then after learning about that, they can explore our garden. We used individual graphs for each laureate to humanize the data and engage interactivity. Also, users can explore each story separately and learn about each laureate.

Author Keywords

Nobel Prize; Data visualization; Scrollytelling; individual graph visualization; storytelling; scientific creativity.

INTRODUCTION

People tend to view high achievers as "finished products" and often overlook the preceding events of a person's life that led to the achievement. This presentation of achievement without context can be alienating to a viewer. To combat this, we present a data visualization that attempts to contextualize the achievements of Nobel laureates alongside information on their career stage, work environments, and personal choices at the time of award receival. In doing so, we hope to provide the viewer with an engaging method to learn about these achievers and gain inspiration.

In this project, we strived to highlight aspects of the laureates that go beyond basic demographics. Visualizations on the demographics of Nobel laureates are readily available, such as on the official webpage of the Nobel prize [7]. Demographic information for a historical award like the Nobel prize tends to spotlight a disheartening pattern of failure to recognize non-male and minority racial groups. For example, the graphic from the Nobel Prize page emphasizes how only 53 women have been awarded out of 923 individual laureates [8]. We recognize that these facts can further alienate an audience, so we chose to focus attention elsewhere.

We believe the attributes we have encoded in our visualization may suggest more insightful trends than demographic information because they can tie more closely to individual characteristics. We include age at the time the laureate received the award to represent experience and maturity when they were recognized. Novelty in a living environment, often from traveling and moving, can be a catalyst for original ideas. We attempt to capture novelty in the environment through a boolean value of whether the laureate was working in a different country from where they were born at the time of award receival. Frequently, highly-skilled scientists are portrayed as lone-wolf geniuses in popular culture. We want to explore the validity of this portrayal by encoding collaboration as an attribute, which we define as a boolean value of whether the laureate shared the prize with other research collaborators.

Our visualization includes the intersection of data from the Nobel Laureate website and dataset from Li [1]. The data from the official webpage includes the laureates' names, affiliations, birth date, and field in which they were awarded the Nobel prize. Li's dataset includes the publication name, year, and journal of all recorded publications for each Physics, Chemistry, and Medicine laureate. From the intersection of Li's dataset with the official Nobel laureate data, we collected information on the number of publications a laureate had before receiving the award to represent persistence and prior recognition in their field.

Because not all laureates are represented in Li's dataset and we wanted to include all Nobel prize fields, we decided to emphasize different aspects of the data using a scrollytelling approach and multiple "flower field" views. We designed each flower to represent the attributes of a laureate in an artistic and visually-stimulating way. Aggregate trends of

the laureates can be determined through filtering and reordering of the flowers, so that each flower field view is essentially a small multiple graph. The individual flower is dense with information, so to keep the user engaged we opted to include a scrollytelling introduction. Through scrollytelling, we frame our research question while also explaining the encodings and data we use.

With our visualization, we encourage the user to explore common conditions and environments that may facilitate academic creativity. This includes examining whether working individually or collaboratively enhances research novelty, or whether top scientists tend to move away from or remain in their birth countries. From these shared attributes, we would like to distill common personality traits of influential thinkers and present them in a compelling way.

RELATED WORK

We include a dataset of publication records for Nobel laureates by Li, J. et al [1]. To create this dataset, Li et al reconstructed publication histories for almost all Nobel laureates in Physics, Chemistry, and Medicine from 1900 to 2016 by combining data collected from CVs, university websites, Wikipedia, and the publication and citation database from Microsoft Academic Graph (MAG) [1]. Their dataset not only allows us to discover novel patterns of productivity and collaboration, but may also help us unearth the fundamental principles underlying creativity and the genesis of scientific breakthroughs.

In "The data behind the Nobel prizes" by Withers, N. et al [2], they focused on chemistry Nobel Prize winners and studied how long laureates waited between publishing their work and winning the prize, how many citations their papers gathered, and where the laureates lived and worked. They used data visualization (Sankey charts) to illustrate where the laureates were born and whether they moved to other countries or not. Also, they illustrated how old the laureates were when they won the award and revealed the gender breakdown between laureates in the field of chemistry. In the last part of the article, the authors illustrated the number of citations the papers gathered and how the most-cited papers got citations . This study uses data visualization to answer the questions and for each question they used different types of visualization to find out the gaps.

In another visualization by Giorgia Lupi [5], she visualized the sleep habits and literary productivity of famous writers. Her data visualization illustrates the patterns of thirty-seven writers for whom wake-up times were available. This became the base data set, around which she set out to quantify and visualize the literary productivity of each author, while also adding the main awards they won and the length of each writer's life to provide additional context. We were inspired by how she tells the story of each individual writer as a distinct graph (Figure 1.). Her goal was to understand whether any correlations exist between famous writers' productivity and the time they used to wake up.

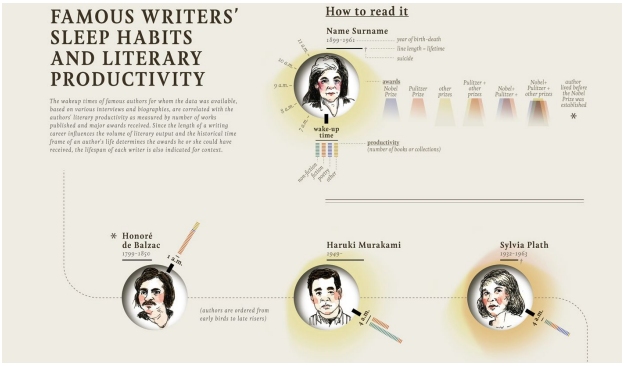

Figure 1. Famous writers' sleep habits by Giorgia Lupi

METHODS

We started our project by brainstorming different ways to visualize the story of laureates in a way to motivate others and have impacts on other researchers. We wanted to see which visualization techniques would help us reach our goals to discover the common conditions and environment that boost scientific creativity, whether showing aggregate visualizations to offer users a big picture and concise takeaway or displaying individual graphs for each laureate to encourage interactivity and user exploration. We created different scenarios to help us to understand which one would be more impactful and informative.

Figure 2. shows our first scenario that is based on aggregate visualization. At the beginning, users see a map with some flags that each one represents a different field. They can learn what is the first award each country won. By clicking on each flag, a graph related to that country shows up which illustrates how many publications that country has during the years and users can have a comparison between different fields. Then, they can click on each category and a word cloud pops up that illustrates the motivations that laureates had in that field and each word is related to one laureate and by selecting the word, users will be taken to the page of that laureate in the Nobel Prize webpage. As you can see, in this scenario, we visualize data in the aggregate

way and users can make some comparison between different fields and countries.

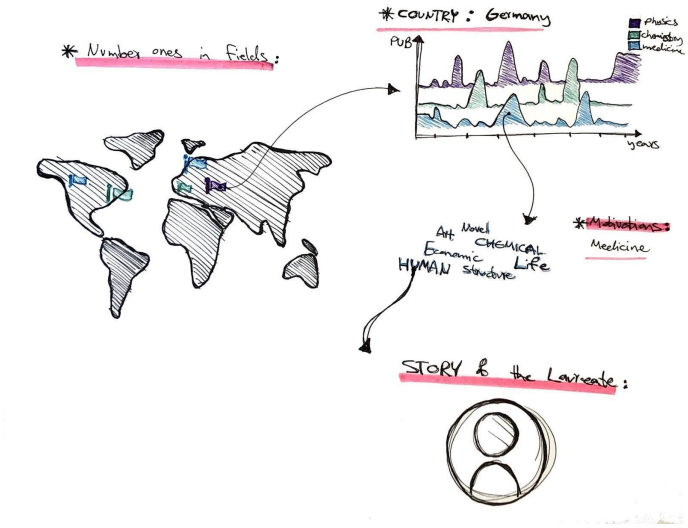

Figure 2: Scenario 1- aggregate visualization

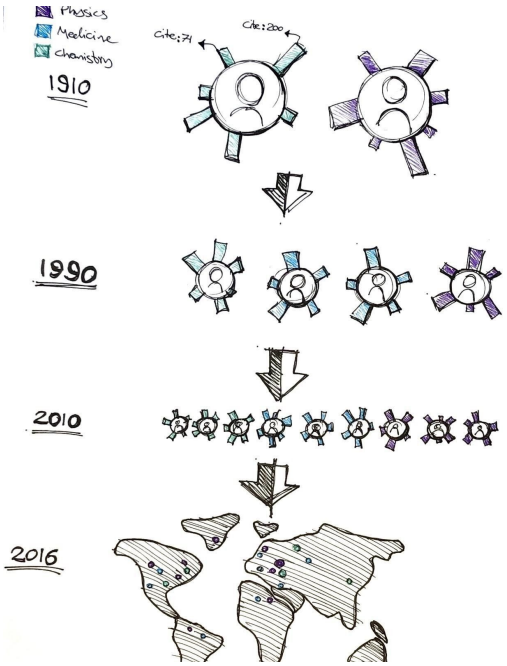

Figure 3: Scenario 2- Scroll telling visualization

We prototyped a strictly aggregate view through development of a chord diagram that represented the flow of laureates between top countries. This visualization proved to be visually cluttered and unclear in its messaging. It was particularly confusing because we realized how unclear it can be to group laureates from the same geographical country together when that same country could have had different names (or been split differently) depending on the time period. We decided not to include this graph in the final application and explore a different approach.

In scenario 2, we created individual graphs for each laureate to humanize the data in a similar way that Giorgia Lupi did. As shown in figure 3, each graph is for an individual laureate, the color encodes their prize field and the bars represent the number of publications they had. The length of the bar also shows the number of citations they received. We combine the individual graphs with scroll telling visualization to enable our users to find out about more laureates in recent years, as they scroll, they can see more recent laureates who won Nobel prize. At the end of the page, each graph goes on the geography map to show the country in which the laureate was born. In this scenario, the audience learn about each Nobel laureates separately and they can learn about their stories.

We combined these two scenarios and included some aggregate components to offer users a big picture and take away like having a word cloud to illustrate the motivations that laureates had in their research field.

We used scroll telling techniques to explain our encodings. We found out that learning about each laureate with individual graphs needs to have different legends and encodings. To explain our legend about the details of graph, we used scroll telling techniques as it is a way to dynamically tell multimedia stories that unfold as you *drum roll* scroll. Scroll telling is a good way to engage the audience as the scroll gives them a sense of control, exploration, and discoverability.

For each graph, we designed flowers to represent each laureate. The color of the flower represents the field that the laureate won the prize. The shape of the petal represents the mobility of laureates whether they moved to another country or stayed in their born country. The number of petals illustrates the number of publications the laureates have, and the leaves show whether the prize was shared with another laureate or if it was a solo research. Lastly, the length of the stem represents the age that the laureate won the prize.

At the end of the page, when users learn about the encodings and legends, they can explore more Noble laureates and go to the flower field. In this page, they can

arrange the garden by filtering the data such as selecting the field (category), the year the laureates won the prize, the age that they won the prize, and the number of publications they have.

By clicking on each flower, a window pops up which shows the summary of the encodings like the figure 4. And then, by clicking on the flower, the user will go to the Noble prize webpage of that Laureate.

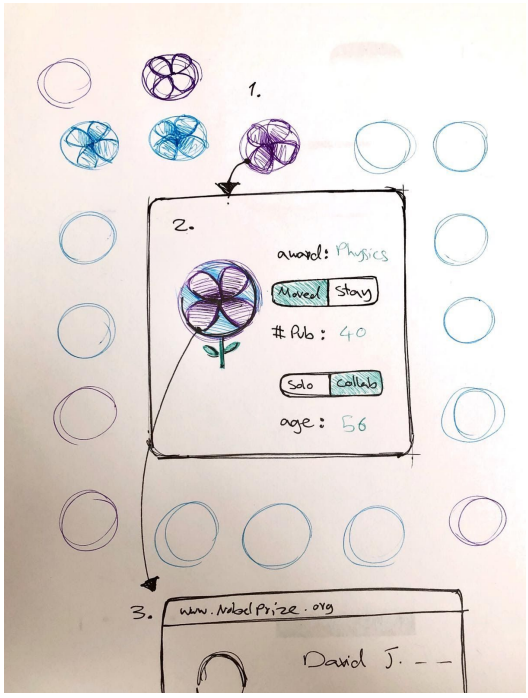

Figure 4. A window example to show the summary of encodings

RESULT

As you can see in figure 5. and the GitHub page, we created our visualization with the Nobel prize dataset that Li, J. et al (2019) gathered. Also, we used the dataset which is available on the Nobel Prize website. The parts we focused in our visualization covers the information of laureates' field, the year they won the prize, whether they moved to another country or stayed in their born country, the number of publications the laureates have, the age they won the prize, and whether the prize is shared with other laureate or if they did their research alone. They can learn about this information and its encodings in the first page by scrolling. As they scroll, the flower will be completed and they can learn about each encoding. You can visit the page and scroll to learn about our flowers:

https://github.mit.edu/pages/6894-sp20/FP-The-Stories-of-High-Achievers/

After learning about the encodings, they can go to the main page which is our garden and the audience can explore all laureates here and they can also arrange the garden by filtering the data (Figure 5.).

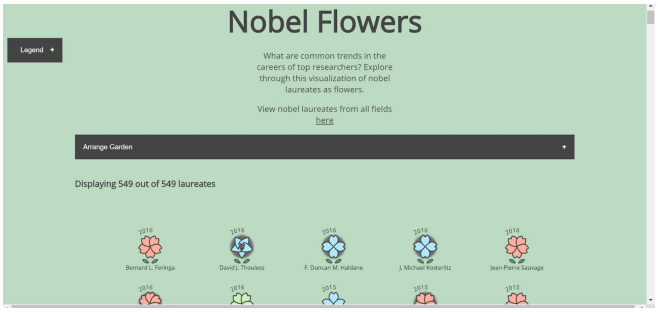

Figure 5. GitHub webpage of Nobel flowers

We created a bar called "Arrange Garden" (Figure 6.) which enables our users to filter the data and flowers to be more organized and to the point. Users can play around with the filtering tabs to explore more about Nobel laureates. They can filter the age laureate won the prize, the number of publications they have, and the year they won the prize by sliding bar. Also, they can filter the field they want to explore and select the order preference.

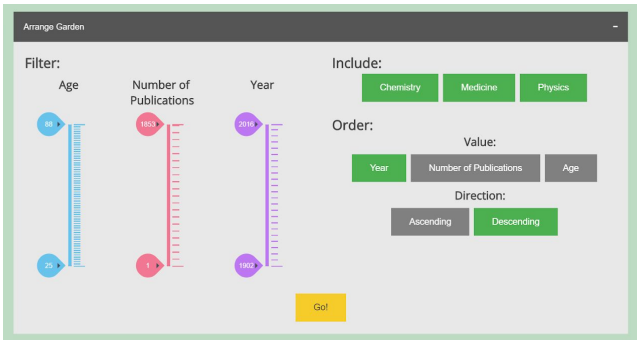

Figure 6. Filtering bar called "Arrange Garden"

By filtering, users can navigate the data more conveniently and less overwhelmingly as there is a huge number of laureates which is hard to explore. They can choose what information they want to explore and what stories they want to observe. Moreover, they can find out in what conditions and environment each laureate won the Nobel prize, whether it was solo or collaborative research, if the laureate moved to another country and in which age they won the prize.

We also added the Legend bar to remind users of our encoding. It is easy to access and as they scroll the flower garden, they can have the legend bar and whenever they select a flower, they can check it with our encodings. The shape of the petal represents the age the laureate won the prize. The color encodes the field of the laureate. Leaves represent whether they collaborate with other researchers or if they work individually. The gray background of flowers represents the laureates' mobility, whether they stayed in their country or if they moved to another country. The number of petal also shows the number of publications laureates have.

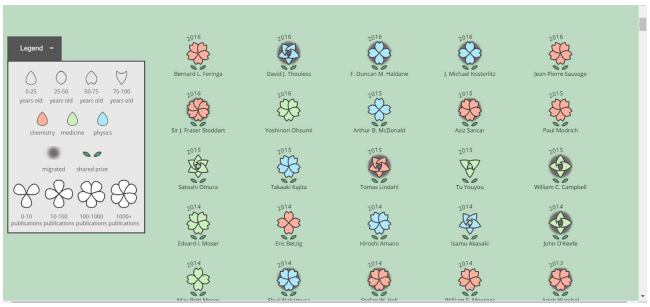

Figure 7. Legend bar of Nobel flowers

After selecting a flower, a window with a summary of information pops up. As shown in figure 8, users can see the flower with information about the field of laureate, whether they migrated or stayed, the number of publications the laureate has, whether the laureate collaborates with other researchers and the laureate's age. For the migration, we designed two boxes called "Migrated" and "Stayed". Based on the laureate information, the box is filled with the color. This also applies for the shared prize, whether they collaborate or not. At the bottom of the box, we give users instruction of how to close the window and how they can go to the Nobel Prize webpage of the laureate.

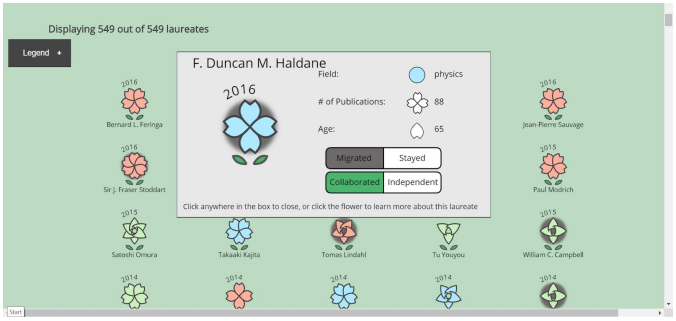

Figure 8. Window of Summary

DISCUSSION

With our visualization and the website, users can learn most of the laureates moved to the other countries. With the background color of the flower, users can find out whether the laureate moved to the other country or not. Also, users can see there are more Nobel prizes in the medicine category than other fields and most laureates won their prize when they are baby boomers and passed their 50s by looking at the shape of the petal.

Although, users can learn about the story of each laureate individually, they cannot do much comparison and this is what we aimed for, to motivate users by telling the story of each Nobel laureate. By looking at each individual graph, they can see their lives and how and when they won the prize, whether they collaborate with others to do their research, if they move to other places, how many publications they have and how old they were when they won the prize.

One of the insights we have found from the visualization is that, before 1970, we had more laureates with the age between 25-35 rather than recent years. And when laureates were in their younger age, they moved less than the time they grew older.

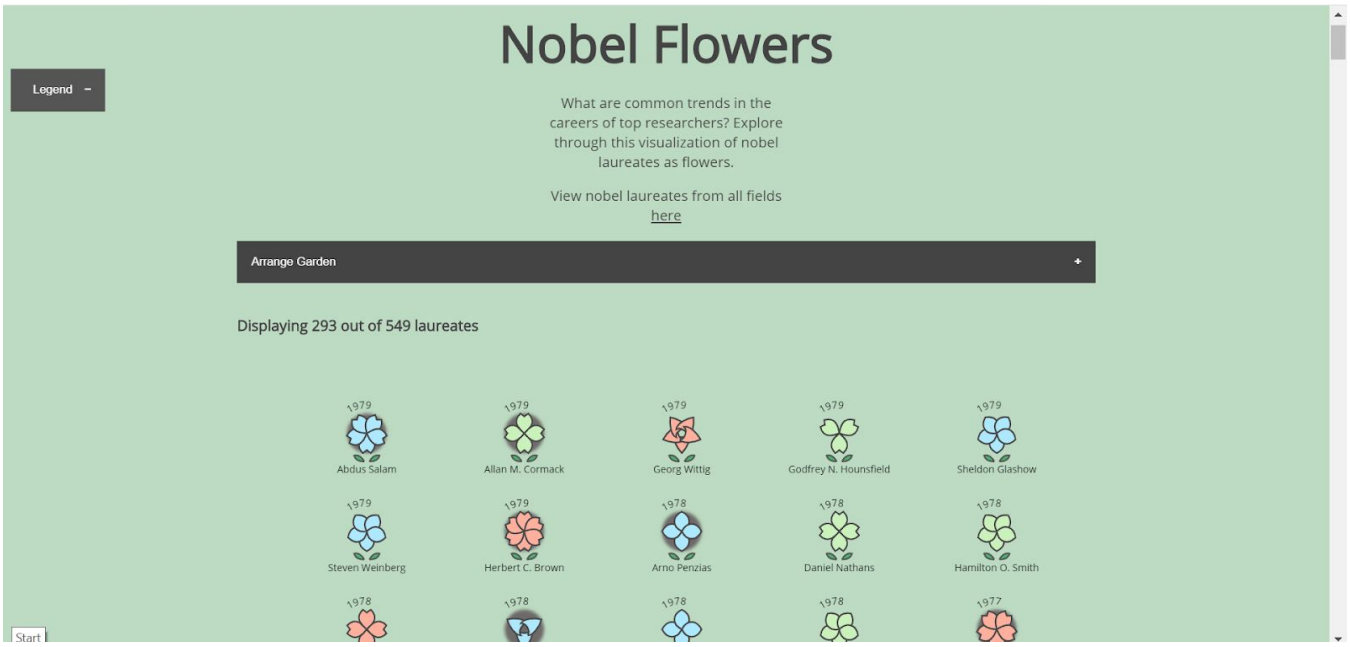

Figure 9. Nobel Flowers

FUTURE WORK

Our visualization can be extended and refined by adding aggregate visualization to enable our users to make a comparison or get a better picture and take away from the visualization. The aggregate visualization can be added as the next level of looking at the flower garden. After filtering and selecting options by users, they can take a look at each graph to learn about the laureates individually and then after navigating through flowers, all flowers can be moved with the animation to the aggregate graphs to show the comparison and complete the story.
Also, for future work, we can add a geography map that illustrates each laureate's born country and by adding animations, like clicking on the laureate, we can show the country the laureate moved to.

Besides, we can visualize the citation that each laureate got by adding the size feature to the flowers, the bigger the petal of flower will be, the more citation the publication has gotten.

REFERENCES

[1] Li, J., Yin, Y., Fortunato, S. and Wang, D. A dataset of publication records for Nobel laureates. *Scientific Data 6*, 1 (2019).

[2] Withers, N., Robinson, P., Valsler, B. and Walter, P. The data behind the Nobel prizes. *Chemistry World*, 2020. https://www.chemistryworld.com/nobel-prize/the-data-behind-the-nobel-prizes/4010453.article?adredir=1&adredir=1&adredir=1&adredir=1&adredir=1&adredir=1.

[3] Microsoft Academic. *Academic.microsoft.com*, 2020. https://academic.microsoft.com/publications.

[4] Visual Data — giorgialupi. *giorgialupi*, 2020. http://giorgialupi.com/lalettura.

[5] Famous Writers' Sleep Habits and Literary Productivity — giorgialupi. *giorgialupi*, 2020. http://giorgialupi.com/famous-writers-sleep-habit-and-productivity.

[6] Nobel Laureates, 1901-Present. *Kaggle.com*, 2020. https://www.kaggle.com/nobelfoundation/nobel-laureates.

[7] The official website of the Nobel Prize - NobelPrize.org. *NobelPrize.org*, 2020. https://www.nobelprize.org/.

[8] Nobel Prize awarded women. *NobelPrize.org*, 2020. https://www.nobelprize.org/prizes/lists/nobel-prize-awarded-women/.